



# **Business Intelligence**

per i Servizi Finanziari

Prepared by: **Simone Somazzi**  
A.A. 2023/2024

# Contents

<b>1</b>	<b>Introduzione alla finanza</b>	<b>4</b>
1.1	Securities . . . . .	4
1.1.1	Bond . . . . .	4
1.1.2	Azioni . . . . .	5
1.2	Asset Return . . . . .	5
1.2.1	Ritorno netto . . . . .	5
1.2.2	Ritorno lordo . . . . .	6
1.2.3	Ritorni con i dividendi . . . . .	6
1.2.4	Ritorni con inflazione . . . . .	6
1.2.5	Ritorni annualizzati . . . . .	7
1.3	Compravendita di azioni . . . . .	7
1.3.1	Order . . . . .	7
1.3.2	Historical Data . . . . .	8
1.3.3	Payoff e Profitto . . . . .	8
1.3.4	Indici . . . . .	8
1.4	Strumenti derivati . . . . .	9
1.4.1	Option . . . . .	9
1.4.2	Future e Forward . . . . .	10
1.5	Volatilità . . . . .	10
<b>2</b>	<b>Portfolio</b>	<b>12</b>
2.1	Gestione del portfolio . . . . .	12
2.1.1	Fondi comuni di investimento e ETF . . . . .	12
2.2	Portfolio Returns . . . . .	13
2.3	Ottimizzazione del portfolio . . . . .	13
2.3.1	Portfolio con risk-free . . . . .	14
2.3.2	Capital Market Pricing Model . . . . .	15
2.4	Selezione Portfolio . . . . .	16

2.4.1	Buy and Hold . . . . .	16
2.4.2	Constant Rebalanced Portfolio . . . . .	16
<b>3</b>	<b>Trading</b>	<b>17</b>
3.1	Posizioni degli investitori e del mercato . . . . .	17
3.1.1	Long position . . . . .	17
3.1.2	Short position . . . . .	17
3.1.3	Mercato Bullish o Bearish? . . . . .	18
3.2	Classi di traders . . . . .	18
3.2.1	Aggressivi, reattivi, passivi o attivi? . . . . .	18
3.3	Attitudine al trading . . . . .	19
3.3.1	Hedger . . . . .	19
3.3.2	Speculativo . . . . .	19
3.3.3	Arbitrageurs . . . . .	19
3.4	Discounted Cash Flow Model . . . . .	19
3.5	Electronic Markets e Limit Order Book . . . . .	20
3.5.1	Perchè è utile analizzarlo? Microprice . . . . .	20
3.5.2	Electronic Market . . . . .	21
3.5.3	Altre tipologie di ordini . . . . .	21
3.5.4	Colocation . . . . .	22
3.5.5	Exchange fees . . . . .	22
<b>4</b>	<b>Forecasting</b>	<b>23</b>
4.1	Modello di serie temporali nella finanza . . . . .	23
4.2	Trend e stagionalità . . . . .	24
4.3	Stazionarietà . . . . .	25
4.4	Indici di performance (errore) . . . . .	25
4.5	Single & Multi step . . . . .	26
4.6	ARMA . . . . .	26
4.6.1	ARIMA (p,d,q) . . . . .	27
4.6.2	ARIMAX(p,d,q) . . . . .	27
4.6.3	SARIMAX (p,d,q)x(P,D,Q) . . . . .	27
4.7	Neural Network . . . . .	27
4.7.1	Percettrone . . . . .	27
4.7.2	Reti neurali . . . . .	28
4.7.3	Backpropagation . . . . .	28
4.7.4	Misure di performance per classificazione . . . . .	28
4.8	K-NN . . . . .	29

4.8.1	Classificazione . . . . .	29
4.8.2	Regressione . . . . .	29
4.8.3	Serie temporali . . . . .	30
4.9	Support Vector Machines . . . . .	31
4.9.1	Metodo Kernel . . . . .	31
4.9.2	Regressione . . . . .	32

# Chapter 1

## Introduzione alla finanza

### 1.1 Securities

Sono strumenti finanziari negoziabili con un determinato valore finanziario. Troviamo principalmente tre tipologie di securities:

- Debt: chiamati anche risk free, sono considerati i più sicuri (es. bond, depositi bancari)
- Equity: azioni di una azienda
- Derivati: il loro valore dipende dal valore di altri fattori (es. future e forward contract, option)

Tutti questi securities sono scambiati nel mercato.

#### 1.1.1 Bond

Vengono chiamati anche obbligazioni, sono contratti tra due parti:

- **Debitore** (issuer): colui che riceve una determinata somma dall'holder
- **Holder**: elargisce una determinata cifra, con la promessa di riceverla indietro insieme agli interessi.

Con **principal** si intende la somma versata dall'holder, mentre i **coupons** sono i successivi interessi elargiti in diversi momenti fissati. La **maturity date** è la data di scadenza del bond.

Il valore del bond dipende quindi dal tempo di maturazione, il tasso di interesse e la frequenza dei pagamenti degli interessi.

Data la somma iniziale  $P_0$  e  $r$  tasso di interesse, con  $n$  numero degli anni e  $m$  numero pagamenti in un anno, posso calcolare il valore finale  $P_n$  con interesse composto:

$$P_n = P_0 \left(1 + \frac{r}{m}\right)^{nm} \quad (1.1)$$

Il payoff è il valore alla fine del titolo (principal + interessi), mentre il profitto è il valore finale risk-adjusted tolto l'investimento iniziale ( $P_n - P_0$ ).

### 1.1.2 Azioni

Le aziende vendono partecipazioni per aumentare il capitale tramite lo stock-exchange market.

La capitalizzazione di una azienda è: numero azioni \* valore singola azione.

Il valore di una azienda quindi varia insieme al valore delle sue azioni.

Ogni possessore di azioni di una azienda è parzialmente proprietario di quest'ultima, in base al quantitativo che possiede.

## 1.2 Asset Return

Definiamo  $P_0$  il prezzo al tempo  $t_0$ , mentre  $P_1$  il prezzo di vendita dell'asset al tempo  $t_1$ . Il ritorno semplice è:

$$R(t_0, t_1) = \frac{P_1 - P_0}{P_0} \quad (1.2)$$

### 1.2.1 Ritorno netto

Il ritorno semplice netto è dato da:

$$R_{t,t-1} = \frac{P_t - P_{t-1}}{P_{t-1}} \quad (1.3)$$

Viene definito semplice in quanto viene considerato un solo periodo.

Invece se abbiamo più periodi da analizzare, possiamo considerare:

$$R_t = \frac{P_t - P_{t-k}}{P_{t-k}} = \frac{P_t}{P_{t-k}} - 1 \quad (1.4)$$

Oppure possiamo moltiplicare i singoli ritorni

$$R_t = \prod_{j=0}^{k-1} (1 + R_{t-j}) - 1 \quad (1.5)$$

### 1.2.2 Ritorno lordo

Il ritorno semplice lordo è dato da:

$$1 + R_{t,t-1} = \frac{P_t}{P_{t-1}} \quad (1.6)$$

Mentre per quanto riguarda più periodi, basta aggiungere +1 ai ritorni netti.

### 1.2.3 Ritorni con i dividendi

Se un asset paga dei dividendi  $D_t$  tra il periodo  $t - 1$  e  $t$ , allora il ritorno netto è:

$$R_t^{total} = \frac{P_t - P_{t-1}}{P_{t-1}} + \frac{D_t}{P_{t-1}} \quad (1.7)$$

Consideriamo che:

- $\frac{P_t - P_{t-1}}{P_{t-1}}$  viene chiamato **capital gain**, ovvero il ritorno dato dalla sola differenza di prezzo iniziale e finale (tassata con aliquote).
- $\frac{D_t}{P_{t-1}}$  viene chiamato **dividend yield**, utilizzato come indicatore del rendimento immediato indipendente dal corso del titolo azionario.

### 1.2.4 Ritorni con inflazione

Otteniamo un vero guadagno se il rendimento sarà maggiore dell'inflazione. Possiamo quindi procedere in due modi diversi. Viene utilizzato  $CPI_t$ , indice del livello del prezzo generale (Consumer Price Index) al tempo  $t$ .

Scontare  $P_t$

$$P_t^{REAL} = \frac{P_t}{CPI_t} \rightarrow R_t^{REAL} = \frac{R_t^{REAL} - R_{t-1}^{REAL}}{R_{t-1}^{REAL}} = \frac{P_t}{P_{t-1}} \frac{CPI_{t-1}}{CPI_t} - 1 \quad (1.8)$$

Utilizzare  $\pi_t$

$$\pi_t = \frac{CPI_t - CPI_{t-1}}{CPI_{t-1}} \quad (1.9)$$

$$R_t^{REAL} = \frac{1 + R_t}{1 + \pi_t} - 1 \quad (1.10)$$

### 1.2.5 Ritorni annualizzati

Per avere una comparazione più omogenea, solitamente i ritorni sono convertiti in ritorni annuali. Dato  $R_m$  il ritorno ogni mese per un anno (12 mesi), il ritorno lordo annuale (CAGR) è:

$$1 + R_A = 1 + R_t(12) = (1 + R_m)^{12} \quad (1.11)$$

## 1.3 Compravendita di azioni

### 1.3.1 Order

Si possono comprare/vendere azioni tramite orders. Esistono principalmente tre differenti order (market, limit e stop).

#### Market Order

L'ordine viene mandato al mercato ed eseguito subito.

Priorità alla velocità rispetto al prezzo, viene eseguito il prima possibile a discapito del rischio di vendere/comprare ad un prezzo leggermente differente.

#### Limit Order

Viene fissato un prezzo (limite guadagni/perdite) a cui eseguire la transazione. Priorità al prezzo rispetto al tempo, so a quanto la vendo ma non quando. I limit order altrui sono visibili (pagando) tramite LOB (limit order book).

#### Stop Order

Viene indicato ad un broker un prezzo (stop price), il quale una volta raggiunto vengo avvisato. Ricevuto l'avviso, posso decidere se vendere o meno. Rispetto al limit order, posso limitare le perdite e decidere i guadagni. Inoltre non sono visibili agli altri come i limit order.



### 1.3.2 Historical Data

Quando vogliamo analizzare determinate azioni, possiamo avere diversi dati:

- Ticker: ID della compagnia (es. TSLA per Tesla)
- Open: prezzo all'apertura del mercato
- Close: prezzo alla chiusura del mercato
- : High: prezzo più alto raggiunto tra apertura e chiusura del mercato
- Low: prezzo più basso raggiunto tra apertura e chiusura del mercato
- Adjusted Close: prezzo di chiusura aggiustato per includere dividendi o altre azioni (es. split)
- Volume: numero di azioni scambiate tra apertura e chiusura del mercato
- Dividend: guadagni dati periodicamente dalle aziende

### 1.3.3 Payoff e Profitto

Supponendo di comprare azioni al tempo  $t_0$  per un prezzo  $S_0$  e rivenderle al tempo  $T$  per un prezzo  $S_T$ , allora il payoff è  $S_T$ . Considerando di ricevere dividendi per un valore  $D_T$  e un tasso di interesse  $r$ , allora il profitto sarà:

$$S_T + D_T - C(S_0)e^{rT} \quad (1.12)$$

### 1.3.4 Indici

Gli indici sono funzioni matematiche che misurano i cambiamenti di prezzo di un gruppo selezionato di azioni (rappresentativo di un mercato o un settore industriale). Viene usato come indicatore di trend di un mercato e del suo stato di salute. Sono molto utilizzati per comparare diversi mercati e per vedere la performance di una azione rispetto al suo mercato.

⚠ Non si può fare direttamente trade di un indice (si usano ETF). Possiamo trovare due tipologie di indici: pesati per prezzo o capitalizzazione.

### **Indice pesato per prezzo**

Tengono conto solo del prezzo della singola azione. Una grande oscillazione di una singola piccola azienda può influenzare molto l'indice.

### **Indice pesato per capitalizzazione**

Considera anche il numero di quote (capitalizzazione = numero azioni \* valore singola azione). Quindi una piccola oscillazione di una grande azienda influenza molto l'indice, mentre oscillazioni di piccole aziende vengono considerate meno.

## **1.4 Strumenti derivati**

### **1.4.1 Option**

Contratto tra due parti dove, tramite il pagamento di una fee, viene riservata la possibilità (no obbligo) di vendere (put) o acquistare (call) un assett entro una futura data (expiration date) ad un prezzo fissato (strike price).

Utilizzate per limitare le perdite, mantenendo la possibilità di non limitare i guadagni. Infatti nel caso peggiore, si perde solo la fee dell'option. Entrambe le parti si assumono un rischio, ovvero l'holder si assume il rischio di pagare le fee, perdendole se non esercita l'opzione. Il writer dall'altra parte rischia di perdere molto, vendendo o acquistato ad un prezzo peggiore rispetto al mercato. Entrambi scommettono su direzioni opposte del mercato.

Possiamo trovare diverse topologie di option:

- Europea: option esercitabile solo durante l'expiration date
- Americana: option esercitabile ogni giorni di trading fino all'expiration date
- Bermuda: tra le precedenti due, possibile esercitare option solo in alcune date fissate fino all'expiration date
- Asiatica: option che valuta l'evoluzione del titolo in termini di media
- Barriera: option associata ad un limite di prezzo (ovvero quando asset attraversa, in salita o in discesa, una certa quotazione), determina l'inizio o il termine della validità

Option europea e americana vengono definite vanilla, essendo semplici e chiare. Le altre vengono definite "esotiche", poichè dipendono da vari fattori.

### Payoff e profitto

Il payoff dipende dall'esercitare o meno l'option.

In vanilla, il profitto per le call option è:  $\max(P_T - k, 0)$ , mentre per le put option è:  $\max(k - P_T, 0)$ .

Per calcolare il profitto finale bisogna tenere conto delle fee e delle commissioni, chiamati prezzo dell'option. Consideriamo questo prezzo come  $C(P_0, T)$  con  $P_0$  prezzo dell'asset all'inizio, insieme con gli eventuali guadagni dati da un risk-free con rendimento  $r$  per il periodo  $\tau = T - t_0$ . Allora il profitto di una call option sarà:

$$\max(P_T - K, 0) - C(P_0, T)e^{\tau r} \quad (1.13)$$

### 1.4.2 Future e Forward

Al contrario delle options, qui si stipula un contratto di obbligo di scambio di asset per uno specifico prezzo (delivery price). Per i forward abbiamo una data specifica nella quale effettuare lo scambio, invece nei future si ha un periodo di tempo (ad esempio un mese). Non è presente una fee, al contrario degli option.

I contratti forward non sono scambiati sul mercato, ma sono contratti fatti tra privati che si accordano su una specifica data e uno specifico prezzo. Invece i future possono essere scambiati e il delivery price è regolato dal mercato.

Solitamente sono utilizzati per le commodities (materie prime).

## 1.5 Volatilità

La volatilità è una misura dell'ampiezza delle fluttuazioni di prezzo di un titolo finanziario in un determinato periodo di tempo. In altre parole, indica quanto il prezzo di un titolo è soggetto a variazioni. Un titolo con alta volatilità è uno il cui prezzo oscilla molto, mentre un titolo con bassa volatilità è uno il cui prezzo è più stabile.

La volatilità è importante per gli investitori perché è un indicatore del rischio di un titolo. Un titolo con alta volatilità è più rischioso di un titolo con bassa

volatilità perché il suo prezzo è più probabile che subisca forti rialzi o ribassi. La volatilità non può essere osservata direttamente, ma viene stimata da modelli matematici che utilizzano i dati storici dei prezzi. Il modello più comunemente utilizzato è il modello di volatilità storica, che calcola la deviazione standard dei rendimenti passati di un titolo per stimare la sua volatilità futura.

# Chapter 2

## Portfolio

Il portfolio è una collezione di una o più securities possedute dall'investitore. Ogni elemento del portfolio viene chiamata positioe. Il valore del portfolio è dato dalla somma dei valori di ogni position, allo stesso modo il profitto del portfolio è dato dalla somma dei singoli profitti.

### 2.1 Gestione del portfolio

Ogni portfolio è strutturato in base alla propensione e tolleranza al rischio del proprietario. La gestione del portfolio si può ridurre a due basiche operazioni:

- Apertura nuove positions: ovvero l'aggiunta nel portfolio di nuove positions con aspettative redditizie
- Chiusura delle positions: vendita delle positions sul mercato, oppure l'aggiunta dell'inverso di una determinata positions, così da annullare ogni eventuale guadagno o perdita.

Queste operazioni sono effettuate in modo dinamico per minimizzare il rischio e massimizzare i profitti.

#### 2.1.1 Fondi comuni di investimento e ETF

Se non si hanno le competenze per gestire personalmente il proprio portfolio, si possono investire i soldi in fondi comuni di investimento, i quali in autonomia gestiranno questi soldi. Possono essere closed-end (numero di azioni del portfolio fisso) o open-end (non ci sono restrizioni sul numero di azioni).

ETF invece sono degli strumenti di investimento collettivi che puntano a replicare i comportamenti degli indici azionari. Questo permette quindi di non investire in singole azioni, ma in un portfolio di azioni appartenenti allo stesso indice che simula l'andamento di quello specifico indice.

## 2.2 Portfolio Returns

Il ritorno di un portafoglio è la somma pesata dei singoli ritorni delle positions. Avendo un portafoglio con  $n$  asset con quote di investimento  $x_i$  tali che  $x_1 + \dots + x_n = 1$ , allora il ritorno sarà:

$$1 + R_{p,t} = \sum_{i=1}^n x_i R_{i,t} \quad (2.1)$$

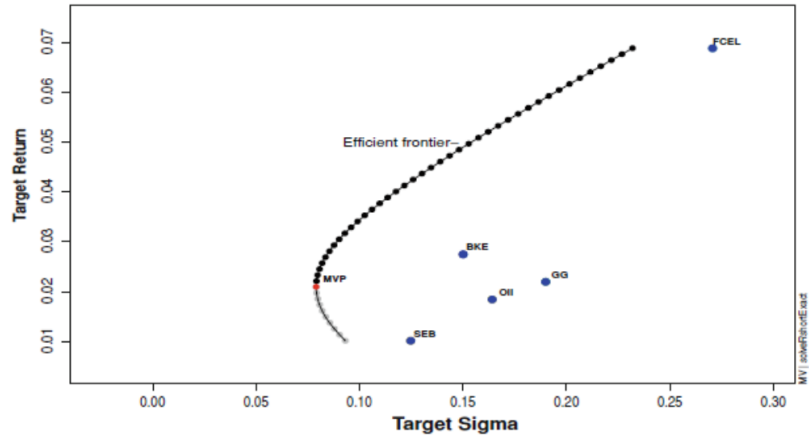
## 2.3 Ottimizzazione del portfolio

Nel 1952 Markowitz ipotizzò una teoria per la costruzione di portfolio basata sul minimizzare il rischio, anche a costo di diminuire i guadagni. Per fare questo, nota come maggiore è la varianza (titoli molto variabili) maggiore è il rischio. La varianza può essere diminuita costruendo il portfolio con asset non correlati tra loro.

Una volta stabilito il ritorno, deciso dall'investitore, si cerca di ottenere con il minimo rischio (minima varianza). Possiamo costruire un problema di ottimizzazione delle quote degli asset dato da:

$$\min w' C w \quad (2.2)$$

soggetto al fatto che  $w' \mu = r^*$  dove  $r^*$  è il rendimento scelto dall'investitore e  $\sum_{i=1}^N w_i = 1$  i vari pesi. Possiamo quindi analizzare il problema graficamente, con l'asse Target return (ritorno) e l'asse Target Sigma (deviazione standard). Come possiamo notare abbiamo una curva, sotto la quale sono presenti infiniti portafogli, mentre sopra non possiamo averne nessuno. La curva viene chiamata frontiera di efficienza. Ogni punto sulla linea rappresenta per ogni target return il minimo valore di deviazione standard (rischio). Possiamo trovare così MPV (minimum variance portfolio), ovvero il portfolio con minor rischio (minor varianza) possibile.



### 2.3.1 Portfolio con risk-free

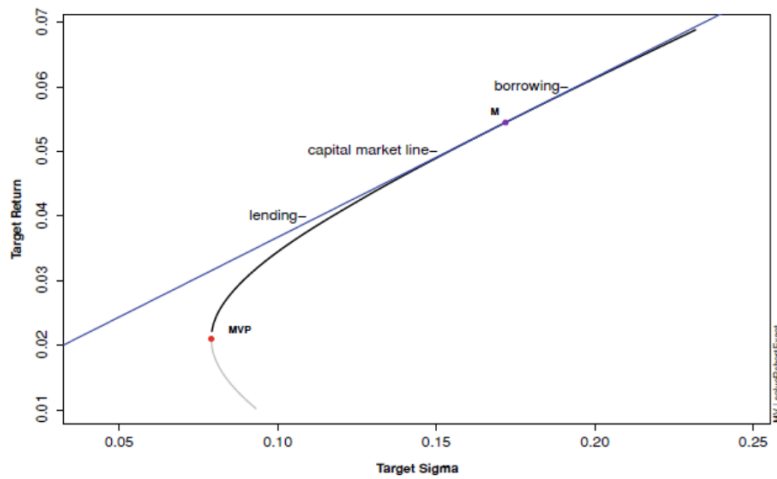
Possiamo anche inserire un asset risk free nel portfolio. Possiamo quindi ipotizzare la divisione dei pesi in  $1 = w_r + w_f$ , dove  $w_f$  è la quota del risk free mentre  $w_r$  il resto. Tra un asset risk-free e uno non la varianza è nulla (nessuna correlazione).

Quindi il ritorno atteso e la deviazione standard sono dati da:

$$\mu_w = w_f r_f + (1 - w_f) \mu_w \quad (2.3)$$

$$\sigma_w = (1 - w_f) \sigma_w \quad (2.4)$$

Unendo il grafico precedente e il grafico di un portfolio con risk free otteniamo:



Possiamo notare come le due linee siano tangenti nel punto  $M$ , il quale rappresenta il miglior portfolio (**Market Portfolio**) per rapporto tra eccesso di ritorno e rischio. Prima di quel punto presto soldi ( $w_f > 0$ ), dopo mi prestano dei soldi ( $w_f < 0$ ). Nel punto  $M$  quindi  $w_f = 0$ .

Con **sharpe ratio** si intende l'eccesso di ritorno rispetto al risk-free (ovvero quanto guadagno in più rispetto a non rischiare). Si calcola quindi sottraendo al rendimento avuto il rendimento senza rischi, dividendo poi per la volatilità (deviazione standard).

### 2.3.2 Capital Market Pricing Model

Sia  $E(R_M)$  il ritorno atteso del Market Portfolio, con  $\sigma_M = std(R_M)$  la sua deviazione standard, e  $r_f$  il ritorno di un risk free in un periodo di tempo  $\tau$ . Considerando  $E(R_i)$  il ritorno atteso di un singolo asset  $i$  con  $\sigma_{iM} = Cov(R_i, R_M)$  la covarianza dei ritorni tra  $R_i$  e  $R_M$ . Allora:

$$E(R_i) = r_f + \beta_i(E(R_M) - r_f) \quad (2.5)$$

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2} = \frac{Cov(R_i, R_M)}{Var(R_M)} = \rho(R_i, R_M) \frac{\sigma_i}{\sigma_M} \quad (2.6)$$

Quindi in breve stabilisce una relazione tra il rendimento di un titolo e la sua rischiosità, misurata con un fattore di rischio  $\beta$ . Questo fattore misura la dipendenza lineare del ritorno dell'asset e del mercato rispetto alla sua volatilità.

$E(r_i) - r_f$  è l'eccesso di ritorno (premio del rischio) di un asset.

#### Coefficiente $\beta$

Usando l'equazione  $\beta_i = \frac{Cov(R_i, R_M)}{Var(R_M)} = \rho(R_i, R_M) \frac{\sigma_i}{\sigma_M}$ , possiamo dare l'interpretazione dei movimenti dell'asset rispetto al mercato.



Valore di $\beta$	Correlazione e volatilità	Interpretazione
$\beta < 0$	$\rho < 0, \frac{\sigma_i}{\sigma_M} > 0$	Asset si muove in direzione opposta rispetto al mercato
$\beta = 0$	$\rho = 0$	Asset e mercato non sono correlati
$0 < \beta < 1$	$\rho > 0, 0 < \frac{\sigma_i}{\sigma_M} \leq \frac{1}{\rho}$	Asset e mercato si muovono insieme. Non so molto sulla volatilità (può essere $>$ o $<$ di quella del mercato)
$\beta > 1$	$\rho > 0, \frac{\sigma_i}{\sigma_M} > 1/\rho > 1$	Asset si muove come il mercato. La volatilità dell'asset è migliore di quella del mercato

## 2.4 Selezione Portfolio

Quando un portfolio devia dal ritorno atteso, bisogna revisionare il portfolio, incrementando o decrementando alcune posizioni (tramite i pesi).

Una strategia di selezione di portfolio per  $n$  periodi di trading è una sequenza di  $n$  decisioni sul vettore  $w^n = (w_1, \dots, w_n)$  dove  $w_t \in W$ . Definiamo  $ALG$  un algoritmo che produce strategie di selezione di portafogli. Il suo output  $S_n(ALG, x_n)$  il fattore di valore di  $ALG$  rispetto alla sequenza di vettori di mercato  $x_n$ .

### 2.4.1 Buy and Hold

Una possibile strategia può essere quella di comporre un portfolio e mantenerlo tale fino alla fine, senza mai modificarlo.

### 2.4.2 Constant Rebalanced Portfolio

Mantiene stabile la distribuzione dei pesi per tutto il periodo di trading. Ad esempio se definisco di avere 60% in obbligazioni e il restante in azioni, al periodo successivo nel caso le azioni siano cresciute molto devo venderne parte e reinvestire i guadagni in obbligazioni per mantenere i pesi 60-40.

# Chapter 3

## Trading

### 3.1 Posizioni degli investitori e del mercato

L'investitore si può trovare in due diverse posizioni rispetto ad una determinata azione: long o short.

In generale la maggior parte degli investitori ha una posizione passiva nel mercato (buy-and-hold), ovvero comprano securities e le mantengono fisse nel portfolio per lungo tempo. Questo perchè la tendenza della maggior parte delle securities è quella di alzarsi con il passare del tempo. Questo però rischia di perdersi eventuali guadagni vendendo ai picchi e comprando successivamente ad un prezzo minore.

#### 3.1.1 Long position

Un investitore che ha una visione rialzista rispetto ad una azione investe su di essa, mettendosi quindi in una posizione long. Quando vende quella azione, si dice che chiude la sua posizione long.

#### 3.1.2 Short position

Un investitore ha una visione pessimista di una azione, si metterà in una posizione short. Questo significa che lui chiederà in prestito  $n$  azioni di quel titolo, vendendole subito dopo e ottenendo  $P_0$ . Se infine tutto andrà come previsto dall'investitore, potrà successivamente ricomprare le stesse  $n$  azioni di quel titolo da ridare, ma ad un prezzo minore  $P_T$ . Così avrà guadagnato la differenza  $P_0 - P_T$  dal calo del valore di una azione.

### **3.1.3 Mercato Bullish o Bearish?**

Un mercato viene definito bullish se tende ad avere un aumento di prezzo, invece è detto bearish se in almeno due mesi il prezzo è sceso del 20%.

Viene definito bull un investitore che vedono i prezzi salire, al contrario viene definito bear.

## **3.2 Classi di traders**

Possiamo classificare gli investitori in tre macrocategorie: fundamental, informed traders, market makers.

### **Fundamental**

Guarda i bilanci e piani dell'azienda prima di investire su di essa. Scelta di investimento è quindi guidata da fondamenti economici non direttamente legati a dinamiche del mercato.

### **Informed Traders**

Cercano di vendere/comprare prima di un eventuale deprezzamento/aumento, valutando il futuro sulla base di fattori esterni al mercato (es. news).

### **Market Makers**

Troviamo qui gli speculatori, che al contrario dei primi due (in particolare rispetto ai fundamental), guarda principalmente alle dinamiche del mercato.

### **3.2.1 Aggressivi, reattivi, passivi o attivi?**

Al contrario del pensiero comune, il quale tende a vedere gli speculatori come attivi e aggressivi, nella realtà dei fatti è esattamente l'opposto. Infatti un mercato composto da soli speculatori sarebbe un mercato prettamente piatto, in quanto loro non muovono il mercato ma si muovono con esso. Coloro che muovono il mercato e quindi li definiamo aggressivi o attivi sono coloro che investono non solo sulla base dell'andamento del mercato ma rispetto a fattori esterni. In particolare quindi i fundamental.

## 3.3 Attitudine al trading

Abbiamo principalmente tre tipologie di attitudine: hedger, speculativo, arbitrageurs.

### 3.3.1 Hedger

Poco propensi al rischio, preferiscono muoversi per minimizzarlo anche a costo di ottenere minori guadagni. Possono avere nel portfolio posizioni contrarie tra loro su diversi securities.

### 3.3.2 Speculativo

Si assumono rischi per poter raggiungere possibili guadagni maggiori.

### 3.3.3 Arbitrageurs

Cercano di guadagnare tramite le differenze di prezzo tra diversi mercati. Ad esempio se a Francoforte il cambio EUR/USD è di 1.25 mentre a New York è 1.3, cercherebbe di comprare a Francoforte e vendere a New York. Questo però non tiene conto di tutti i costi di transizione che ci sono, inoltre i prezzi tendono ad allinearsi quasi subito, lasciando la possibilità di effettuare questi scambi praticamente nulla.

## 3.4 Discounted Cash Flow Model

Il valore di una azione dipende dalla possibilità che offre di generare guadagni nel futuro. Possiamo quindi stimare il present value  $S_0$  di una azione (tot anni in avanti) sulla base del futuro prezzo  $S_t$  e i dividendi previsti  $D_t$ , al tasso di interesse  $r$  come:

$$S_0 = \sum_{t=1}^T \frac{D_t}{(1+r)^t} + \frac{S_T}{(1+r)^T} \quad (3.1)$$

Possiamo ottenere  $r$  con:

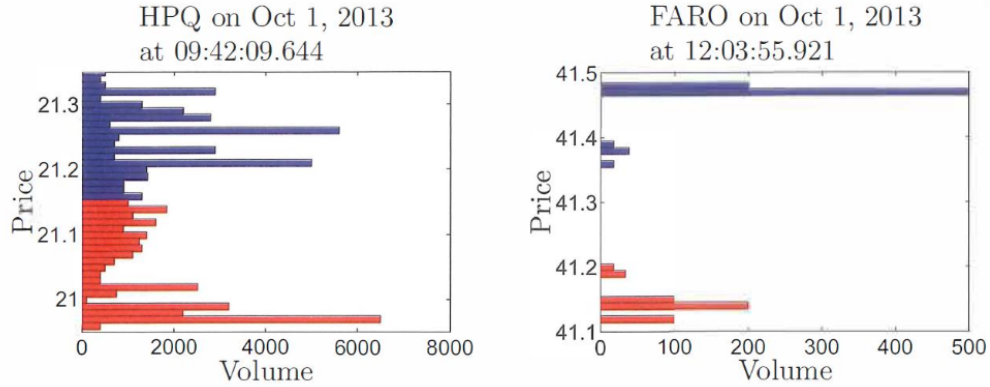
$$r = \frac{S_1 + D_1}{S_0} - 1 \quad (3.2)$$

## 3.5 Electronic Markets e Limit Order Book

Possiamo definire il LOB come una griglia discreta dei prezzi (livelli). La differenza tra un livello e il successivo è chiamato **tick**, il suo valore cambia da mercato a mercato (in US equivale ad un cents per le azioni che valgono più di un dollaro). Contiene tutti i limit order di vendita e acquisto con i relativi prezzi e numero di azioni.

La differenza tra il miglior prezzo ask ( $P_t^a$ ) e bid ( $P_t^b$ ) è chiamato **quoted spread**. Maggiore è la differenza, minore è liquido il titolo, ovvero avvengono meno scambi (essendoci tanto divario tra domanda e offerta).

Viene definito come **midprice** la media aritmetica tra best ask e best bid.



### 3.5.1 Perchè è utile analizzarlo? Microprice

Molto utile analizzare il LOB poichè sono visibili le posizioni degli altri investitori, con i relativi prezzi e volumi. Si può quindi studiare gli scambi e prevedere eventuali incrementi o decrementi dei prezzi in base al volume di ask e bid.

#### Microprice

Viene definito come prezzo tarato in base al volume di bid ( $V_t^b$ ) e ask ( $V_t^a$ ):

$$Microprice_t = \frac{V_t^b}{V_t^b + V_t^a} P_t^a + \frac{V_t^a}{V_t^b + V_t^a} P_t^b \quad (3.3)$$

Questo può essere utilizzato come anticipatore dei prezzi, infatti in molti casi il prezzo futuro dipende anche dal volume di domanda e offerta, oltre che dal divario di prezzo tra loro.

### 3.5.2 Electronic Market

Grazie al LOB, possiamo abbinare market order in entrata ai limit order. Generalmente i mercati danno priorità ai market order rispetto ai limit order.

Viene utilizzata di solito la politica **price-time priority**. Quando arriva un market order prendo in considerazione il best-price per soddisfarlo, e tra tutte le azioni con il best-price prendo le più vecchie. Se non sono presenti abbastanza azioni nel livello di best-price, possiamo applicare diverse politiche.

#### Walking the book

In questo caso viene preso il secondo best-price con le sue azioni. Se non bastano neanche le azioni di questo secondo livello, quelle restanti dopo aver consumato (comprato) queste azioni si passa al livello successivo migliore, così via fino a quando non ho comprato il numero desiderato di azioni.

#### re-routing

Se necessito di altre azioni oltre a quelle già presenti nel best-price, prendo il numero di azioni che mi mancano, una percentuale faccio walking the book, mentre il restante viene comprato al primo best-price in un altro mercato.

#### Immediate-or-Cancel

In questo caso l'ordine viene parzialmente soddisfatto con solo azioni al best-price, il restante ordine (richiesta che eccede) viene cancellato.

### 3.5.3 Altre tipologie di ordini

Possiamo trovare altre tipologie di ordini per poter attuare diverse strategie:

- **Day Order**: sono ordini che avvengono durante il mercato e sono estesi prima dell'apertura o dopo la chiusura.

- **Non-routable:** orders che non possono essere reindirizzati
- **Pegged:** se il midprice o best-price cambia, ordine viene spostato senza costi di cancellazione (obv costo maggiore all'inizio)
- **Hidden:** non si possono visualizzare le loro quantità (non si può calcolare così microprice).
- **Iceberg:** mostrano una quantità inferiore rispetto a quella reale. Quando quella fittizia viene consumata, viene introdotta la restante.
- **Fill or Kill:** Consumi tutto al prezzo miglior price o cancelli.
- **Good-Till-Time:** order con una fine temporale prestabilita.
- **Discretionary:** ordini che nascondono prezzi più aggressivi.

### 3.5.4 Colocation

Una delle componenti più importanti nel trading è la velocità. Per questo si possono noleggiare dei server vicini ai computer che effettuano i matching degli order. Questi server devono essere al pari, anche a livello hardware, quindi tutti i cavi devono essere lunghi uguali e il network uguale per tutti a livello di prestazioni e velocità.

### 3.5.5 Exchange fees

Tutti gli scambi effettuati non sono esenti da fee, le quali variano da mercato a mercato, in particolare abbiamo due principali differenti approci.

#### Maker-taker system

Aiuta chi inserisce LO (aggiunge liquidità al mercato) con minori fee o con premi, rispetto a chi effettua MO (togliendo liquidità al mercato) con fee più alte.

#### Taker-market system

Al contrario qui abbiamo vantaggi per chi effettua MO rispetto a chi inserisce LO.

# Chapter 4

## Forecasting

Analisi delle serie temporali per previsioni future.

### Assunzione di linearità

Data una serie temporale  $X_{t+h}, \dots, X_t, \dots, X_{t-p}$  dove  $X_t$  è il valore di ieri e la serie segue una distribuzione gaussiana, allora esiste una combinazione lineare tale che:

$$E(X_{t+h}|X_t, \dots, X_{t-p}) = \alpha_0 X_t + \dots + \alpha_p X_{t-p} \quad (4.1)$$

Dove  $E$  è il valore dell'asset al tempo  $X_{t+h}$ , il quale quindi dipende dalla combinazione lineare dei suoi precedenti valori.

Devo riuscire a costruire un predittore avente i migliori coefficienti possibili (minimo errore)  $\alpha_k$  tramite la regressione lineare.

### 4.1 Modello di serie temporali nella finanza

Ogni componente della serie  $X_t$  è composta da una parte deterministica  $H_t$  e un rumore  $Y_t$ , quindi  $X_t = H_t + Y_t$ .

La parte più difficile è stimare il rumore, calcolato come:

$$Y_t = E(Y_t|F_{t-1}) + \alpha_t \quad (4.2)$$

$$Var(Y_t|F_{t-1}) = E(\alpha_t^2|F_{t-1}) = \sigma_t^2 \quad (4.3)$$

dove  $F_{t-1}$  rappresenta il set di informazioni disponibile a un tempo  $t - 1$ ,  $E(Y_t|F_{t-1})$  è la media condizionata,  $\alpha_t$  è lo shock stocastico (o innovazione) e



si presume abbia media condizionale nulla, e quindi la varianza condizionale. Possiamo ottenere diversi modelli per la previsione, in particolare quelli più semplici sono basati solo su serie temporali con assunzione di linearità. Tra questi troviamo **ARMA** e **ARCH**, il primo mantiene la varianza fissa, nel secondo varia in base a  $F_{t-1}$ .

Possiamo poi costruire modelli più complicati, basati su funzioni non lineari, come NN e SVD.

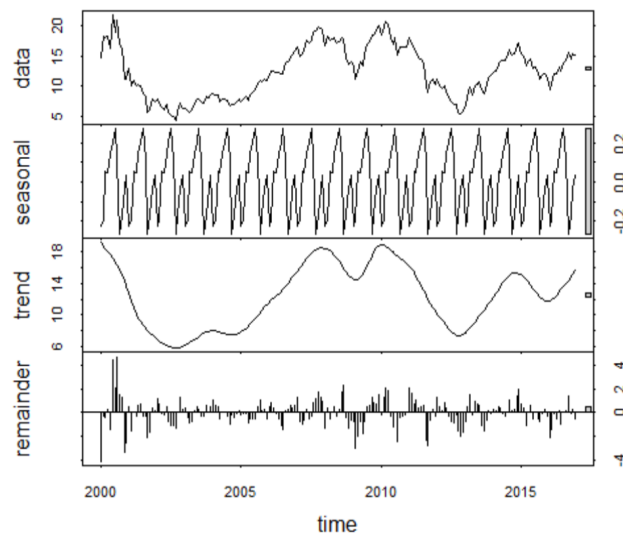
## 4.2 Trend e stagionalità

Nella formula  $X_t = H_t + Y_t$  la componente deterministica  $H_t$  comprende trend, stagionalità e ciclicità. Possiamo quindi vederla come:

$$H_t = m_t + s_t \rightarrow X_t = m_t + s_t + Y_t \quad (4.4)$$

In particolare:

- $m_t$  è la componente di **trend**, una funzione che cambia lentamente.
- $s_t$  è la componente di **stagionalità**, una funzione periodica.



Una volta che stabilisco la componente deterministica, posso eliminarla dai dati per ottenere il solo rumore  $Y_t = X_t - H_t$ .

## 4.3 Stazionarietà

Nella maggior parte dei casi le serie temporali non sono stazionarie. Ci sono diversi approcci per stabilirlo, uno di quello più comuni è il test statistico ADF. Si possono poi applicare delle trasformazioni per poter rendere le serie stazionarie (ad esempio considerare i ritorni logaritmici della serie).

### Test ADF

Basato sulla regressione:

$$X_t = \alpha + \beta t + \varphi X_{t-1} + \epsilon_t + \sum_{i=1}^{p-1} \theta_i \Delta X_{t-i} \quad (4.5)$$

Dove  $\alpha$  è una costante,  $\beta$  il coefficiente del trend,  $p$  è il numero di ritardi, scelto secondo criterio statistico (AIC o BIC).

Se  $\alpha = \beta = 0$ , allora abbiamo una random walk (ovvero i valori seguono un movimento casuale), mentre se  $\alpha \neq 0$  e  $\beta = 0$  abbiamo una random walk with drift.

Il risultato del test AIC dipende dal valore di  $\varphi$ , ovvero se  $\varphi = 0$  la serie non è stazionaria, se invece  $\varphi < 0$  lo è.

## 4.4 Indici di performance (errore)

Possiamo utilizzare diverse tipologie di errore per i vari modelli:

- **MSE** (Mean Squared Error):  $\frac{1}{m} \sum_{t=1}^m (f_t - a_t)^2$ , ottimo come loss, ma unità di misura al quadrato.
- **RMSE** (Root Mean Squared Error):  $\sqrt{\frac{1}{m} \sum_{t=1}^m (f_t - a_t)^2}$ , per visionare l'errore mentre modello usa MSE.
- **MAE** (Mean Absolute Error):  $\frac{1}{m} \sum_{t=1}^m |f_t - a_t|$
- **MAPE** (Mean Absolute Percentual Error):  $\frac{1}{m} \sum_{t=1}^m \frac{|f_t - a_t|}{a_t}$

## 4.5 Single & Multi step

Data una serie temporale:

$$X_{t-p:t} = \{x_i\}_{i=t-p,\dots,t} \quad (4.6)$$

possiamo avere tre differenti task:

- Previsione del **singolo valore successivo**:  $x_{t+1} = f(X_{t-p:t})$
- Previsione del **singolo valore  $h > 0$  steps dopo**:  $x_{t+h} = f(X_{t-p:t})$
- Previsione dei valori di una serie su una **finestra temporale** con larghezza  $h$ :  $X_{t+1:t+h} = f(X_{t-p:t})$

I primi due sono single-step, ovvero calcolo un valore, mentre l'ultimo è definito multi-step, ovvero bisogna calcolare più valori consecutivi. Il calcolo di questi valori consecutivi può avvenire in due modalità:

- Multi-Output: tutti i valori  $X_{t+1:t+h}$  sono calcolati insieme (one-shot)
- Ricorsivo: single-step riprodotto ricorsivamente usando anche le predizioni precedenti per predire lo step successivo.

## 4.6 ARMA

### AR(p)

Autoregressivo di ordine  $p$ , è la combinazione lineare dei valori precedenti più  $W_t$  (rumore):

$$X_t = W_t + \sum_{i=1}^p \phi_i X_{t-i} \quad (4.7)$$

### MA(q)

Media mobile di ordine  $q$  degli errori passati, sempre considerando  $W_t$  (rumore):

$$X_t = W_t + \sum_{i=1}^q \theta_i W_{t-i} \quad (4.8)$$

## ARMA(p,q)

Autoregressivo e media mobile di ordine  $p$  e  $q$ . Unione delle prime due, ovvero considera i valori precedenti fino a  $p$  periodi prima e i residui (errori) fino a  $q$  periodi prima:

$$X_t = W_t + \sum_{i=1}^p \phi_i X_{t-i} + \sum_{i=1}^q \theta_i W_{t-i} \quad (4.9)$$

### 4.6.1 ARIMA (p,d,q)

Deriva dal modello ARMA, vengono applicate le differenze di ordine  $d$ , numero necessario per poter rendere stazionarie le serie temporali. Quindi un modello ARIMA è analogo a ARMA ma utilizzato sulle serie delle differenze di ordine  $d$  della serie e non sulla serie originale. Se però la serie presenta una forte stagionalità, i modelli ARIMA non sono adeguati.

### 4.6.2 ARIMAX(p,d,q)

Derivano dal modello ARIMA, ma con l'aggiunta di input exogenous (esogeno), ovvero esterni alla serie temporali (es. sentiment).

### 4.6.3 SARIMAX (p,d,q)x(P,D,Q)

Derivano dal modello ARIMAX, per gestire componente stagionale. Abbiamo quindi quattro nuovi parametri stagionali  $(P,D,Q,x)$ ,  $x$  è il numero di step temporali in ciascun periodo stagionale.

## 4.7 Neural Network

### 4.7.1 Percettrone

Il percettrone è composto da connessioni input (a cui è associato un peso), operazioni interne applicate agli input e connessioni output. All'interno viene effettuata una somma pesata degli input, il risultato poi è passato ad una funzione di attivazione, che può essere lineare o non. I pesi vengono chiamati parametri, e in base all'errore ottenuto considerando il dato reale rispetto all'output del percettrone, vengono modificati per minimizzare l'errore.

		True Class	
		Positive	Negative
Predicated Class	Positive	TP	FP
	Negative	FN	TN

## 4.7.2 Reti neurali

Purtroppo un singolo perceptrone non è in grado di ottenere buoni risultati in task complesse, per questo vengono utilizzati più perceptroni insieme, raggruppati in layer consecutivi. L'output di un perceptrone è così l'input del successivo. Il primo layer viene chiamato input layer, l'ultimo output layer e i restanti hidden layers.

## 4.7.3 Backpropagation

Per modificare i valori dei parametri (pesi) si utilizza il metodo backpropagation, ovvero tramite un gradiente, ottenuto dall'errore dell'output e dall'iperparametro **learning rate** (per controllare quando un errore deve valere sulla modifica dei pesi). Se stiamo utilizzando una rete neurale, i pesi vengono modificati a partire dagli ultimi layer fino ad arrivare all'input layer.

## 4.7.4 Misure di performance per classificazione

Viene utilizzata la matrice di confusione, ovvero: Possiamo da questa matrice ottenere diverse misure di performance:

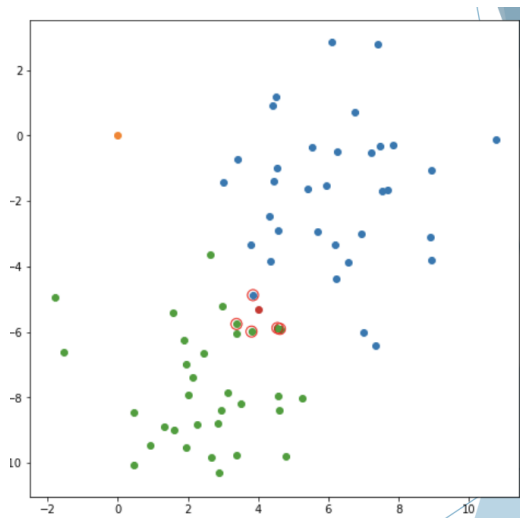
- Accuracy:  $\frac{TP+TN}{FP+FN}$
- Precision:  $\frac{TP}{TP+FP}$
- Recall:  $\frac{TP}{TP+FN}$

## 4.8 K-NN

Data in input una istanza, l'output associato è predetto in base alle più vicine  $k$  istanze del dataset. Si può utilizzare per la classificazione, per la regressione e per le serie temporali.

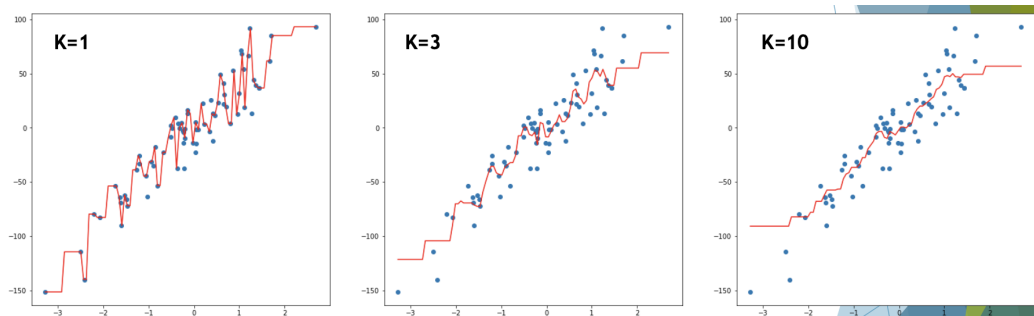
### 4.8.1 Classificazione

Tramite una funzione distanza calcolo i primi  $k$  vicini, la classe dell'istanza in input sarà la classe con più presenza nei  $k$  vicini.



### 4.8.2 Regressione

Simile al procedimento per la classificazione, viene fatta la predizione del valore in base al valore dei primi  $k$  vicini (ad esempio la loro media). Non è necessaria la linearità dei dati, infatti si adatta comunque.



I punti principali sono:

- Più alto il numero di  $k$ , maggiore sarà la morbidezza della curva predetta.
- Posso usare qualsiasi funzione di distanza
- Posso pesare diversamente ogni vicino
- Semplice implementarlo, funziona anche con non lineari, non ci sono ne parametri ne fase di training
- Su grandi dataset diventa lento, dovendo stimare la distanza di ogni punto.

### 4.8.3 Serie temporali

I dati sono serie temporali, si basa sulla similarità per comparare diverse serie temporali. Abbiamo due iperparametri principali:  $k$  e  $p$  (numero di osservazioni passate che vogliamo considerare per generare forecasting).

Cerca le  $k$  sottosequenze nella serie che sono più 'vicine' alla serie temporale data in input  $X_{t-p:t} = \{x_i\}_{i=t-p, \dots, t}$

Utilizza poi il futuro di ognuna delle  $k$  sottosequenze simili per generare la previsione:

- **Single-step:**  $x_{t+h} = g(\{x_{t_i+h}\}_{i=1, \dots, k})$ , dove  $t_i$  è la fine della  $i$ -esima sottosequenza tra le  $k$  più vicine, mentre  $g()$  è una funzione per calcolare la media (pesata o no).
- **Multi-step:** vogliamo predire  $X_{t+1:t+h}$ , per cui  $g()$  non può essere una funzione scalare.

## 4.9 Support Vector Machines

L'obiettivo di un algoritmo SVM è quello di trovare un iperpiano che separi, al miglior grado possibile, i punti di dati di una classe da quelli di un'altra classe. Per “migliore” si intende l'iperpiano che ha il margine maggiore tra le due classi. Questo obiettivo viene raggiunto utilizzando una parte minimale del dataset di allenamento, i cosiddetti vettori di supporto (ovvero i dati più vicini all'iperpiano divisore, i dati più difficili da classificare).

Al contrario della classificazione lineare, tramite SVM è possibile effettuare classificazioni non lineari tramite il metodo kernel.

### 4.9.1 Metodo Kernel

L'idea di base è quella che istanze non separabili linearmente possono essere separate in uno spazio differente, dove i punti sono mappati diversamente. Questo cambiamento di spazio per ogni punto sarebbe molto pesante dal punto di vista computazionale, per cui si usano i 'kernel tricks'. Esso permette di calcolare il prodotto scalare tra vettori di dati trasformati senza la necessità di effettuare esplicitamente la trasformazione. Troviamo diverse tipologia di kernel functions:

- **Polinomiale**
- **Radial Basis Function**

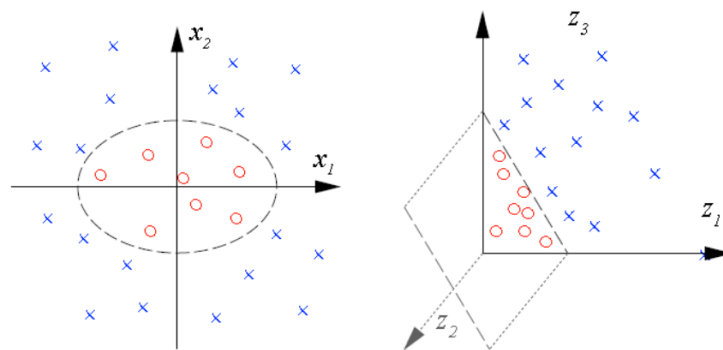


Figure 4.1: Polinomiale



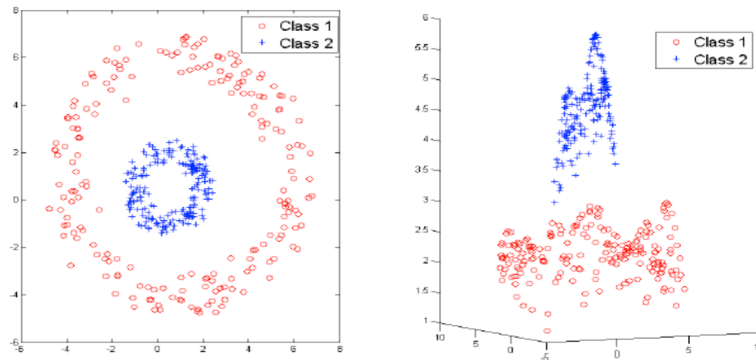
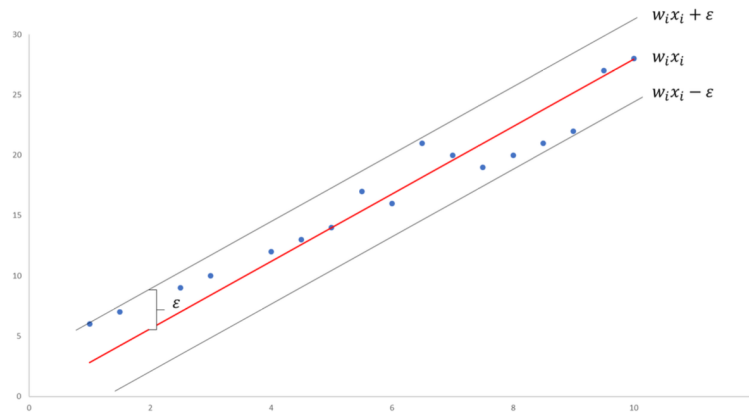


Figure 4.2: RBF

### 4.9.2 Regression

Può essere utilizzato anche per la regressione, cerca di trovare l'iperpiano di regressione che massimizza il margine di supporto intorno ai punti di addestramento, con una penalizzazione per i punti di addestramento che cadono al di fuori del margine.



Si possono utilizzare anche qui i 'kernel trick', inoltre aggiungo l'iperparametro  $C$ .

### Iperparametro $C$

Controlla il trade-off tra la minimizzazione dell'errore di classificazione e la massimizzazione del margine decisionale.

Un valore più grande di  $C$  implica una penalità maggiore per le violazioni del margine e può portare a modelli SVM più complessi e adattati ai dati di addestramento. Un valore più piccolo di  $C$  consente una maggiore tolleranza per le violazioni del margine, che potrebbe portare a modelli più semplici e con una migliore generalizzazione.

Posso trovare il miglior valore per  $C$  tramite una greedy search.